if, instead of μ being known, there was known only some more general functional relationship between the mean and the variance, the maximum likelihood estimates of these will generally involve both statistics, and not generally be sufficient. The joint sufficiency, however, of $\bar{x}$ and s^2 implies equally the joint sufficiency of any two independent functions of n, $\bar{x}$, and s^2, if such functions can be regarded as estimates at all.

In fact, I think the distinction you are drawing is one without an essential difference, one's choice of an unbiassed estimate being arbitrary in the sense that it is only justifiable by the use to which the estimate is intended to be put.

E.C. Fieller to Fisher: 21 *November* 1946

A copy of the tenth edition of *Statistical Methods*, which we have had on order for some months, arrived here last week, and one of my colleagues promptly pointed out the identity of the argument of the new section 26.2 with the remarks that I published in 1940 and 1944 on the fiducial limits of a ratio. I am enclosing a copy of the latter paper;[1] the former appeared in Vol. VII of the Statistical Society Supplement.[2]

The fact that you do not refer to either of these papers makes me fear that I have inadvertently overlooked an earlier publication of your own; could you let me know, so that I may make due acknowledgment as soon as possible? I could most rapidly do so by adding a footnote to a paper that I have just sent off for publication.

[1] Fieller, E.C. (1944). A fundamental formula in the statistics of biological assay, and some applications. *Q. J. Pharm. Pharmacol.* **17**, 117–23.
[2] Fieller, E.C. (1940). The biological standardization of insulin. *J. R. Statist. Soc., Suppl.* **7**, 1–54.

Fisher to E.C. Fieller: 23 *November* 1946

Thanks for your letter. If you like I could refer to your note in subsequent editions. The general method of substituting values of pivotal quantities such as t and χ^2 at chosen levels of significance in order to make equivalent tests of parameters contained implicitly in these quantities has, I suppose, been creeping into use for a long while. Certainly in the case which you treat I had discussed the quadratic equations to which your method leads with C.I. Bliss, when he was working on toxicological problems at the Galton Laboratory. If anyone has anticipated your note in this application it would be he, but I do not know that he has done so. I seem to remember him using the intercepts of a given ordinate with the Working and Hotelling hyperbola.

E.C. Fieller to Fisher: 29 *January* 1947

I am sorry to have taken so long to answer your letter of November 23rd — as opportunity offered I have been looking up Bliss's publications, and some of them have proved rather inaccessible.

Bliss certainly inverted the Working–Hotelling argument in his 1935 paper, but I doubt whether he then realised the general argument, since he went on using Gaddum's Pearsonian formula in comparative assays until 1944. Paulson gave the fiducial distribution of the ratio of two correlated normal estimates in 1942, but so did I in 1940; what led me to it then was needing to find fiducial limits in an insulin assay after doing a covariance correction; but I tried to write the result in a general form. I would of course take it as a compliment if in subsequent editions of *Statistical Methods* you referred to my note — I agree that the basic idea of equating the significance level of t to the expression (once one has found it) involving the unknown α is a standard one; I think I first learnt it from §62 of *Design of Experiments*. . . .

E.C. Fieller to Fisher: 10 *January* 1954

I believe that Irwin sent you a few weeks ago a draft of a paper by Monica Creasy, in which she seemed to have arrived at an anomalous result concerning the fiducial distribution of a ratio, and mentioned that we were contemplating devoting the March meeting of the R.S.S. Research Section to a discussion of that and related topics.

Of course, we would not want to have such a meeting if you felt that we were likely to confuse rather than clarify the issue, so I am enclosing a revised version of Miss Creasy's paper,[1] and about the first third of the draft that I am preparing to precede her effort. . . . Miss Creasy and I would both very much appreciate your views on our drafts. She is clearly answering a different question from the usual one, but I am not sure that she has yet succeeded in expressing clearly what the difference is; and I would like to be told, if you think that I am committing any howlers.

Miss Creasy is genuinely worried, in particular, to find an appropriate name for her limits; she feels that the title in her draft, 'fiducial distribution limits', is not the right one, but does not like the only alternative, 'apparent limits', that I have been able to suggest.

If you are able to spare the time, what we would both like to do would be to come up to Cambridge one day within the next week or two, and ask you to sort out our minds for us! May we do that?

P.S. I am checking with Chester Bliss that I am not taking his name in vain in my introduction.

[1] Eventually published as Creasy, M. (1954). Limits for the ratio of means. *J. R. Statist. Soc.* B **16**, 186–94.

Fisher to E.C. Fieller: 11 *January* 1954

Thanks for sending me the two long screeds by Miss Creasy and yourself which I am afraid I have not time to examine with all the care they doubtless deserve.

I hope Miss Creasy will be persuaded no longer to associate my name with the method which in fact I have never used, for I was always satisfied with Bliss's approach, and was heartily glad to see the last of the awkward forms discussed by Geary, with their infinities appearing within the range of integration.

I wonder who put Miss Creasy on to such an unrewarding job.

Of course I should be very happy if you both came up to have lunch with me but I scarcely think I shall do anything to clarify your minds.

E.C. Fieller to Fisher: 13 *January* 1954

Thank you for yours of the 11th, but I am left wondering what to suggest to the Research Section Committee next week. Perhaps we can have a word about that before or after to-morrow's Council meeting; I am inclined to complete my screed anyway, and propose to the Committee that they should consider it and something of Miss Creasy's for our May meeting, possibly cancelling the one planned for March. I shall still hope to persuade you to read my draft, when I have finished it; for one thing, I would not want to use the word 'fiducial' in any new context without first referring to you.

I understand that Miss Creasy embarked on her investigation of her own free will, regarding it as an exercise in the manipulation of fiducial probabilities that would lead her to the usual result. When she arrived at a different one, she should of course have been told to find the flaw in her argument. Instead, she was unfortunately encouraged by her then colleagues at Oxford to believe that her new limits were somehow more fiducial than the real ones. I think that she is now disabused of that idea, but on the face of it her mathematics leaves the reader wondering why an approach that is valid when we are considering the difference of two means should not also be applicable to a discussion of their quotient. The point seems to me to be an interesting and not unimportant one, and George Barnard told me last night that he thinks that he has an explanation.[1] . . .

[1] See Barnard's letter of 14 June 1954 (p. 10).

Fisher to E.C. Fieller: 14 *January* 1954

Thanks for your letter. Of course I did not wish to interfere with your arrangements. I was only giving my personal reaction.

I imagine that the discrepancy may well be due to the omission, in one treatment or the other, of some latent assumption; e.g. it would seem to me that the T_1, T_2 treatment would be appropriate when x and y are known to be independently distributed, but that your treatment is more general and makes no such assumption. In inductive reasoning generality has not the same meaning as when reasoning is deductive for in the latter we are under no obligation to take into account data which we do not use. Consequently, the more general treatment in a sense includes and supersedes the less general.

This, however, is not at all true in inductive reasoning where we ought to expect a different result whenever we ignore any datum of the problem.

D.J. Finney to Fisher: 28 *May* 1946

I have recently been reading and discussing your reply to E.B. Wilson on the four-fold table (*Science*, *94*, pp. 210–11 [*CP* 183]). While I agree entirely with your solution of the problem which you state, I feel that there is still a difficulty unexplained; I should be very interested to hear your views on this, as the point is one likely to occur quite frequently.

You state the problem: 'Of six treated mice five have died and one lived, while of six controlled mice one has died and five lived.' Clearly you are then only interested in the possibility that treatment improves the survival rate and therefore base your test on one tail of the distribution. The probability is determined as in your letter [to *Science*]. But Wilson's original statement of the problem referred not to treated and control mice but to mice receiving virus A or virus B; he is therefore presumably interested in the possibility either that B is more harmful than A or that A is more harmful than B—i.e. in both tails of distribution. Since he has six mice in each group, the two tails are symmetrical and he can obtain his probability by doubling the result for one tail. What happens if he has eight mice in one group, six in the other? Say:

	Died	Lived	Total
A	5	3	8
B	1	5	6
Total	6	8	14

How is he to test the null hypothesis that A and B are equally harmful, while considering deviations from equality in either direction? Simply to double the total probability for

$$\begin{Bmatrix} 5 & 3 \\ 1 & 5 \end{Bmatrix} \text{and} \begin{Bmatrix} 6 & 2 \\ 0 & 6 \end{Bmatrix}$$

scarcely seems appropriate, as it does not correspond to any discrete subdivision of cases at the other tail such as

$$\begin{Bmatrix} 1 & 7 \\ 5 & 1 \end{Bmatrix} \text{and} \begin{Bmatrix} 0 & 8 \\ 6 & 0 \end{Bmatrix}.$$

Nor does there appear to me any obvious reason for calculating the probabilities for the two most extreme configurations at the other tail

(keeping marginal totals unaltered) and adding their total to the appropriate probability for the tail at which the observations occur.

Am I missing something very simple here? I cannot remember having seen this problem discussed, and should be grateful for your views. I hope that I am in time to catch you before you go to the States.

Fisher to D.J. Finney: 31 *May* 1946

Thanks for your letter. It is a good problem, but I believe I can defend the simple solution of doubling the total probability, not because it corresponds to any discrete subdivision of cases of the other tail, but because it corresponds with halving the probability, supposedly chosen in advance, with which the one observed is to be compared. That is to say, one may decide in advance that if the probability is less than one in forty in either direction then we shall consider if, pending further investigation, the viruses are not pathologically equivalent.

How does this strike you?[1]

[1] This letter and the last part of Finney's letter of 28 May 1946 were published in Yates, F. (1984). Tests of significance for 2 × 2 contingency tables. *J. R. Statist. Soc.* A 147, 426–63.

D.J. Finney to Fisher: 30 *November* 1949

Suppose that samples from two normal distributions of equal mean give sample means $\bar{x}_1, \bar{x}_2$ and variances of those means s_1^2, s_2^2. Then

$$\bar{x}_1 - \mu = s_1 t_1,$$
$$\bar{x}_2 - \mu = s_2 t_2.$$

If we wish to estimate μ from the combined evidence of the two samples, and to assign fiducial limits to our estimate, I think we can legitimately take

$$\bar{x} = \tfrac{1}{2}(\bar{x}_1 + \bar{x}_2)$$

and use the Fisher–Behrens distribution to give the limits.

We might hope, however, that a weighted mean, $\bar{x}_0$, defined by

$$\bar{x}_0\{(1/s_1^2) + (1/s_2^2)\} = (\bar{x}_1/s_1^2) + (\bar{x}_2/s_2^2)$$

would be more precise. Now

$$\bar{x}_0 - \mu = \{(t_1/s_1) + (t_2/s_2)\}\{(1/s_1^2) + (1/s_2^2)\}^{-1}.$$

Do you see any objection to the use of the Fisher–Behrens argument on this equation? We can integrate

$$(t_1/s_1) + (t_2/s_2)$$

over the region for which it exceeds $d\{(1/s_1^2) + (1/s_2^2)\}^{\frac{1}{2}}$ and it seems to me that the logic should be just as good as in the case you have discussed. I am primarily anxious to be reassured that I am not falling into heresy, before I proceed further. However, I have looked at the next step, in the light of your 1941 paper (*Ann. Eugen. 11*, p. 141 [*CP* 181]). If we modify your transformation at the top of p. 152, writing

$$t_1 = x\cos\theta + y\sin\theta,$$
$$t_2 = -x\sin\theta + y\cos\theta,$$

and

$$s_1/s_2 = \cot\theta,$$

I think that the same integral is obtained. If this is so, the Fisher–Sukhatme tables complete the solution to the problem, but with a modified definition of θ. If I have slipped on this last step, presumably the method you used could be adapted to the formation of new tables.

This idea seems very elementary. Is it new? Its usefulness is clear. For example, it gives the complete solution to the estimation problem in an incomplete block experiment, when a final set of means is formed by combination of inter- and intra-block estimates. A weighted mean of k constituents could be tackled in the same way, but $(k-1)$ ancillary θ-statistics would be needed, and the preparation of tables would be an unpleasant task.

Fisher to D.J. Finney: 2 *December* 1949

Thanks for your letter. I have not checked the algebra, as I have full confidence that you have done it right. But as the problem is one I have run into before, perhaps I might make a few comments.

I think you assume in your first sentences that the samples are of equal size, but will not wish to limit the discussion to this special case.

In general the simultaneous likelihood of the two samples for varying values of μ when maximised for variations of σ_1 and σ_2 does, with confounded perversity, necessarily involve the difference between the means as a minute source of information for both of these unknowns. So far as I can judge, any competently practical solution will ignore this source of information and use s_1 and s_2 only, just as you have done. In fact I have used, and I believe Frank Yates concurs in using, the Behrens–Sukhatme tables for just this purpose. It might be worth your while to see, however, whether Frank has formed any firm opinion on the matter. . . .

D.J. Finney to Fisher: 13 *September* 1954

A few days ago we were discussing Fisher–Behrens, and a remark of yours has sent me to re-read your presentation of fiducial inference in *Annals of Eugenics*, *6*, (1935) pp. 391–398 [*CP* 125]. May I trouble you with one or two questions that worry me, not in what you say but in what is left unsaid?

Of course I appreciate that the fiducial argument is strictly based on

sufficient statistics, and realize that ambiguities are introduced by trying to base the same type of probability statements on statistics that do not utilize all the information contained in the data. Nevertheless, are there not situations in which a sufficient statistic cannot be obtained from the whole data, and only by discarding some information is it possible or practicable to make any probability statement? Must we then deny the validity of any probability statement, or can we use some kind of approximate fiducial distribution?

Three simple examples occur to me:—

(1) If we have a random sample of n from a normal distribution of unknown mean and variance, we can in the usual way make fiducial statements about these two parameters. Suppose now we have one extra random individual for whom all we know is that his value of x exceeds a particular numerical value (e.g. x is the total leaf area of a plant, and some leaves of one plant were accidentally lost). We can apply a maximum likelihood process to estimate the parameters from the $(n+1)$ observations, but I scarcely think that we shall have sufficient statistics. If we know the awkward observation to be truly a random one, are we really being illogical as well as slightly wasteful (only very slightly if n is large) by choosing to neglect it and base our fiducial statements upon the n observations as though the $(n+1)^{\text{th}}$ never existed?

(2) In one of the more complex lattice designs, various types of inter-block mean squares have expectations linearly compounded of inter- and intra-block variances. In addition, there will be an intra-block mean square estimating the intra-block variance alone. Yates and others have shown how to combine the mean squares so as to estimate the two variances. However, if we merely wish to make statements about the intra-block variance and there are rather few degrees of freedom for other mean squares, we might well be content to use the intra-block mean square alone and base fiducial statements on it with the aid of the χ^2 distribution. Are these invalidated by the existence of additional information in other mean squares?

(3) On pp. 296–7 of the *Journal of the Royal Statistical Society*, *B 12* (1950) I suggested the use of the Behrens–Fisher distribution for obtaining a fiducial distribution of a weighted mean of two items, where the items came from distributions with different standard deviations. You had pointed out that the difference between my two items in fact contained a small amount of extra information on the two variances, though I think you will agree that it is not very easily utilized. If x_1, x_2 are means of n_1, n_2 observations, both from distributions with mean μ, and we define

$$\bar{x} = wx_1 + (1-w)x_2,$$

then

$$(x_1-\mu)/(s_1/\sqrt{n_1}) \text{ and } (x_2-\mu)/(s_2/\sqrt{n_2})$$

are both t variates, and my proposal (probably not the first time this has been suggested) was to use Behrens–Fisher on $(\bar{x} - \mu)$ with variance $(w^2s_1^2/n_1) + \{(1 - w)^2s_2^2/n_2\}$. Since $(x_1 - x_2)$ contains information, must we consider this as an entirely heretical use of fiducial argument? I have for a long time been wanting to ask your further advice on this.

Please don't imagine that I am trying to catch you out on this! — except in so far as I am caught myself. For if we refuse to allow any validity to probability arguments along these lines we seem to be denying ourselves access to conclusions that must in some sense be 'very nearly correct', since the information neglected is small. In my example (1), it seems hard that the $(n + 1)$th observation should make us less able to draw a useful conclusion from the data than we were before it was obtained (or than we should be if the experimenter quite legitimately chose not to report it). On the other hand, if we allow an approximate fiducial argument in circumstances where we cannot find a sufficient statistic, or do not want to be bothered with a little extra information, can we deny all validity to the argument based on s' on pp. 392–393 of your paper? Or, perhaps, of more practical importance, if R is the range in a sample of n, the distribution of a 'pseudo-t'

$$(\bar{x} - \mu)/R$$

is known. If computing time is much more precious than time spent in making observations, why should not inferences on the probable position of μ be based on this distribution?

I suspect that I am being very stupid at some point, and I hope that you will put me right. Nothing is further from my mind than to quarrel with the fiducial argument, but I do find a real difficulty here. There's no hurry, as this has been on my mind for a long time.

Fisher to D.J. Finney: 14 *September* 1954

I believe you started writing to me about my old paper on the fiducial argument and then went off to write the rest of the letter about something else. I do not suppose you expect me to deny what I have so often asserted, namely that there are situations in which a sufficient statistic cannot be obtained from the whole data, or the more general proposition that I must have reiterated repeatedly, that the concept of mathematical probability is inadequate to express the nature and extent of our uncertainty in the face of certain types of observational material, while in all cases the concept of mathematical likelihood will supply very helpful guidance, if we are prepared to give up our irrational urge to express ourselves only in terms of mathematical probability.

However, most of your letter is not really on this point but on another. Is there any logical difference relevant to the interpretation of a body of data between such statements as:

(a) I did not get any measurements of that spider because it escaped.
(b) I do not know where I wrote those measurements down, so they are not included in my tabulation.
(c) I could not clearly read what the figures were.
(d) I discarded the observations as untrustworthy.
(e) I could not be bothered to do the calculations all over again.
(f) I had tables based on the range so could ignore intermediate values.
(g) I have found a very good way of calculating my regression line which is sixty times faster than Berkson's,[1] not always quite so accurate (though quite adequately so in most cases!), namely by guessing the number which seems plausible.

I suppose it will be agreed that in all cases a more accurate result could be obtained by giving more care and attention to the problem, and that no line can be drawn between different levels of precision save for specific technological purposes where the amount of tolerance allowable is understood. The point of principle which influences me is not one of greater or less precision, but of misrepresentation to the consumer of the alleged statistical result. He is, as it seems to me, deceived if he is given a value supposedly based on the body of material available but really utilising only a fraction of the information in that body. In fact, the important fact is concealed from him that a more scrupulous examination of the data might have given materially different results.

I conceive that it is possible only by giving students the opportunity of making rather fine distinctions in the logic of the subject, that they can learn to recognize the difference between honest and dishonest work in statistical practice.

In section (3) you seem to be asking my advice but, of course, I was very glad when you showed me what you were doing with the weighted mean from samples of populations with different variances. I pointed out that some information had been neglected, but I do not think I ever encouraged anyone to try to recover what was lost.

Though it may seem hard, it is undoubtedly true that additional information may not only require an adjustment of our conclusions, but also prevent us from making new conclusions in the same form. Of course, in my opinion, we can always ignore any information we like, provided we are scrupulous enough to say what we have ignored. I think a good statistician, in such cases, will wish to present the data in more than one way, these different ways having, if possible, different imperfections.

Some day I hope you will write to me about fiducial inference!

[1] cf. *CP* 256, p. 138.

D.J. Finney to Fisher: 22 *September* 1954

I am sorry if I wrote about the 'infiducial' argument instead of the fiducial. But am I really as far from the topics of your 1935 paper as you suggest? Of course, the first paragraph of your reply is familiar stuff, but I appreciate this very concise statement. My difficulty is that I cannot see the logical difference between the seven statements you tabulate, at least in respect of any clear dichotomy, yet apparently some lead to fiducial arguments and others do not.

May I vary the statements slightly?

(a) I measure a sample of 20 spiders. From the measurements, I can assign a fiducial distribution to the population mean, by use of the standard fiducial argument in its simplest form.

(b) I intend to measure 20 spiders, but 3 escape. Apart from the risk that escapers are non-random, I can proceed to form a fiducial distribution for the mean exactly as in (a), but with 16 d.f. instead of 19 d.f.

(c) I measure 20 spiders, but accidentally lose a random 3 of my records. I can proceed as in (b), as I am using all the information available to me in the form of sufficient statistics, so that the fiducial argument still applies.

(d) I measure 20 spiders, but wilfully and frivolously discard a random set of 3 measurements. I can make exactly the same calculations as in (c), but is my inference about the population mean still to be classed as fiducial? I have been foolish, but scarcely criminal in thus discarding information.[1] I can state quite honestly what was done, and, after my initial folly, I have used sufficient statistics for all the information that remains.

(e) I measure 20 spiders and use the known distribution of the ratio of '(sample mean − population mean)' to 'sample range' in order to make statements about the population mean. The ratio can be manipulated algebraically as is the t value in the fiducial argument, but the argument is no longer fiducial – at least so I have always understood and so I read your 1935 paper. Yet, if (d) is a fiducial argument, why do we make so severe a logical distinction between discarding 3/20 of the observations and discarding a fraction of the information that cannot be associated with a particular subset of observations?

(f) I read a paper in which X reports that he measured 20 spiders, but he records only the mean and the range of the sample. What type of inference can I make about the population mean?

Now it seems to me that (a), (b), (c) correctly employ the fiducial argument, whereas (e) and (f) do not, the confusion I have built up being a consequence of identifying the technical and colloquial usages of the word *information*, with (d) as a borderline quibble that could be classified only after a pedantic insistence on exact definition. This appears to be in line with your 1935 paper, and in particular with your comments on the use of the 'mean error', but surely there *is* then some logical difference between the statements tabulated in your recent letter: (a) and (g) there are at least as distinct as (a) and (f) above. Moreover, the use of the likelihood concept seems to support this.

I want to avoid what you describe as an irrational urge to express myself only in terms of mathematical probability, but I think it is not as irrational to wish to be able to recognize the situations in which such expression is possible and proper.

I am most grateful to you for your last helpful letter, and especially for the valuable statements in the first and last (main) paragraphs. I hope that you may be moved to comment further on this. I know your own reluctance, but I still maintain that you would do a great service to statisticians by producing a new book on inductive inference.

[1] Fisher has written in the margin, 'If stated, no one is misled.'

Fisher to D.J. Finney: 27 *September* 1954

I judge from the letter that you agreed with me that though in all cases information has been lost, yet that the statistician as a professional with his own ethics has to distinguish between cases in which it is lost by his own professional fault, and cases in which he or others have lost it, but not by using inferior statistical methods. The distinction is largely one aimed at discharging the professional duty to giving such services as the client has a right to expect, and, therefore, the fault is largely obviated in case (d) by saying what one has done, or indeed in case (e) if the report makes clear that more accurate and probably stronger statistical statements would have been made by a more sedulous analysis.

In my opinion it is quite arbitrary but very convenient to confine such phrases as 'the fiducial probability' and 'the fiducial distribution', in which the definite article is used, to cases in which these probability statements are unique and as good as could possibly be made from the same data. I would not think it unreasonable to point out that in view of certain difficulties of the data rendering a unique fiducial inference impracticable, yet a fiducial distribution could be inferred from certain portions of the data, or ignoring certain aspects of it, which might be better than nothing. I recur to the point, however, that the mere wasting of information has indeed an economic importance like the wasting of soap, but that what we need to make clear to our colleagues and students is the difference between honest and conscientious work on the one hand, and dishonest and semi-fraudulent work on the other.[1] . . .

[1] For the remainder of this letter, see p. 255.

D.J. Finney to Fisher: 12 *March* 1955

. . . I have a little paradox for you, arising from Section 74 of *The Design of*

Experiments. There you show that if x has a sampling variance s^2 based on n degrees of freedom the precision of x is

$$(n+1)/\{(n+3)s^2\},$$

very properly less than it would have been if the variance of x were *known* to be s^2. Suppose now different investigators each obtain an x and an s^2 from the same population. What is the average of their assessments of precision? Since

$$\mathrm{E}(1/s^2) = n/\{(n-2)\sigma^2\},$$

where σ^2 is the true variance, this seems to be

$$n(n+1)/\{(n+3)(n-2)\sigma^2\} > 1/\sigma^2.$$

Thus on an average the precision is greater than if σ^2 were known!

I suppose the catch is the usual kind of thing in a fiducial argument, namely that the fiducial distribution of μ given s does not relate to sampling from a distribution with fixed σ^2. But what is the right answer to the problem for sampling with fixed σ^2? For example, x is to be the main effect of factor A in a factorial experiment on 4 blocks of 8 plots. If I include only 3 factors, I have s^2 based on 21 d.f. By confounding in a fourth factor, I reduce my d.f. for error to 14, while presumably keeping the same true variance σ^2. In deciding which design to adopt, I want to balance the advantages of studying more factors against the loss of precision to be expected on each of the original main effects and interactions. Ought I to take the ratio of the values of $(n+1)/(n+3)$ to represent the relative precision in respect of A? This gives a reasonable figure, but seems to ignore the fact that the two estimates of s^2 have different distributions.

Fisher to D.J. Finney: 15 *March* 1955

. . . I may well have to think further about your paradox, but you might like to know my first reaction.

I have lately been thinking that although I have, for my own part, always made clear that I introduced the word 'fiducial' to qualify a form of argument by which a true or classical probability statement can be arrived at, yet that its use as an adjective to qualify the word 'probability' has been used, not without malice, to imply that there were two or more kinds of probability used in mathematics, and that the fiducial argument led to a special or peculiar kind called fiducial probability.

I have recently been thinking a little about the semantics of this word. What seems to be implied whenever it is used is a distinction of three levels of knowledge, (a) in which nothing whatever is known about some supposed value, (c) in which the exact value is known, and (b) in which a statement can be made in terms of the concept of mathematical probability, in which the case of a stochastic, or random variable, about which a complete set of

probability statements can be made, is typical.

Any definition of what is the logical content of a probability statement, therefore, must make plain two things, (1) wherein our factual knowledge differs from complete ignorance, and (2) wherein it differs from perfect knowledge. In respect of the first, it is usual to say that the probability of a member of aggregate A being also a member of B refers to a situation in which the aggregates are measurable, and that fraction of set A which belongs also to B occupies a fraction p of the whole set A, this fraction being termed the mathematical probability. Of course the 'measure' in such a definition is only a mathematical abstraction, or generalized term intended to cover the relationship made familiar by such concepts as frequency in statistics, area in plain geometry, or mass in a mechanical situation.

In mathematical books, however, I do not find any complementary statement which clearly distinguishes the concept of probability from that of complete exact knowledge, and I suggest that the players of the eighteenth century, and the manufacturers of appliances for gambling at the present time, are, or were, perfectly familiar with the second requirement which has been omitted from the mathematical statement; that the gambler throwing a die needs to know not only that there is a set of possible throws, and that one sixth of this set are aces, the aggregate of possible throws being in this respect heterogeneous and with a well defined measure ratio, but that, *and this I take to be the necessary and sufficient condition*, none of the sub-sets of the entire set in which the proportion of aces is different from one sixth can be recognized before the die is cast. Of course there always will be such sub-sets, and a claimant to the faculty of pre-cognition claims to be able to recognize a sub-set, namely that in which he foresees that an ace will be thrown, and that the sub-set so recognized does have a proportion of aces differing from one sixth. To such a pre-cognizer, therefore, the probability is not one sixth, but is greater or less than one sixth according to which set he recognizes the throw to belong to. The same would be true of defects in the gambler's apparatus, or corrupt practices in their use, so that the basis on which the mathematician can interest himself in the theory of probability is that of perfect apparatus fairly used, in which, in fact, the second aspect of the definition of probability, namely that which defines the subjective ignorance needed for a probability statement, is justifiable.

This second side of the definition of mathematical probability is not necessary in the formal pure mathematics in which the concept may be used, in which knowledge and ignorance are not at all distinguished, but is likely to be necessary whenever the concept is applied to the real world in which the distinction between what we know and what we do not know is relevant.

It makes plain, I think, rather simply what has puzzled people about the fiducial argument in which the mathematical steps are as follows:

(i) It is proved that the probability of the inequality $T < \theta$ is equal to some known quantity P for all values of the statistic T obtained by random sampling from the population characterized by the parameter θ. The inequality has then a mathematical probability of being realized in the reference set of all possible samples from a single population.

(ii) It is proved that this relationship holds whatever may be the value of θ and, therefore, that the reference set can be enlarged to include all samples from all populations of this type.

This probability statement will subsume the whole of the information available from the observations on two conditions, namely that there is no given *a priori* distribution of θ, and secondly, that the statistic T is exhaustive or sufficient. Without these two conditions, we cannot be in a position to make a probability statement about a parameter, for we should have not used all that is known about it. If a distribution *a priori* had been in the data, we could have applied Bayes' argument, and utilized the whole of the information, to obtain the frequency with which the probability statement is realized in the sub-set having T constant. In the absence of such information *a priori*, and of more informative statistics, there is no means of recognizing any sub-set, whether characterized by T, or by any other means, having a probability different from that of the entire set specified. The frequency ratio in the entire set, therefore, is the probability of the inequality being realized in any particular case, in exactly the same sense as the frequency in the entire set of future throws with a die gives the probability applicable to any particular throw in view; i.e. it is a statement of classical probability arrived at by an argument with which the eighteenth century writers were unfamiliar, and in which the probability statement would have to be modified if we were to change our pre-suppositions and add new knowledge, such as that provided by a Bayesian distribution *a priori*.

I do not believe that this need to specify specific ignorance as well as specific knowledge affects your problem with s^2. My formula involving the ratio $(n + 1)/(n + 3)$ was, of course, only calculated as that appropriate to the amount of information about the parameter supplied by a single value called x with variance estimated from n d.f. If your different investigators are certain that their x and s^2 come from the same population, they would, I suppose, pool their estimates of s. But it might be that they thought their true variances might be conceivably in any ratio however high, as in the case of Behrens' samples, and the discussion would then come down to the precision of the estimate based on the combined evidence of the two samples, in which certainly the difference between the two means must be involved as it throws light on the precision of both sets of measurements.

Probably, however, you are not discussing the combination of observations, though you seem to be doing so in considering $1/s^2$ as though it were going to be used as a weight in some such combination.

I hope I have not been too long-winded about probability. I am concerned at the amount of confused rubbish, which seems to have got itself written, about what I should only regard as a rigorous and useful line of argument.

One trouble was that in 1930 [*CP* 84] I was not at all clear that information *a priori* must be positively excluded, although I was primarily writing about such cases.

D.J. Finney to Fisher: 22 *March* 1955

. . . Your discussion of 'fiducial' is most interesting. Let me first admit that, entirely through ignorance and without malice, I for a long time thought that you distinguished fiducial probability from classical probability, and have no doubt myself propagated this heresy. How I first gained the idea, I do not know, whether from false instruction or misreading, but evidently I failed to find explicit statements to the contrary to correct my folly. I mention this in no spirit of controversy, for I think I now see this matter correctly, but only as evidence that malice is not a pre-requisite of misunderstanding! Call it ignorance, folly, or stupidity if you will, but the effect is the same. You say yourself that your own ideas have become clearer since 1930, and I very much wish that you would now produce in book form an integrated account of the theory of estimation. This would be of inestimable value to the plodders like myself, and ought to stimulate much work in the right direction. You may reply that if anyone wants this, he can do it for himself, but there is a great difference between what you as originator of so much of the theory could do by reinterpreting 30 years of work, and what someone else could do by merely re-writing your papers into one long book. I know I have tried you on this before, but I still hope to persuade you.

I like your (a), (b), (c) distinction on pp. 1–2. Should not the 'logical content of a probability statement' also be qualified by specification of other information known or assumed in the course of the analysis? For example, in estimating the median height of Abominable Snow Men from the small sample of 10 collected by the 1959 Himalayan Expedition we might assume (and for a sample of less abominable Englishmen might think we knew) that

(1) The distribution of individuals is normal with S.D. 0.3 feet;
or (2) The distribution of individuals is normal, parameters unknown;
or (3) The distribution looks normal but has a finite range whose limits are at any rate not wider than 0.5 feet, 20 feet;
or (4) The distribution is unimodal, continuous, and always positive;
or (5) The distribution is continuous;
or (6) Nothing.

However, I am not brought much nearer to explaining away my paradox. The idea of combining information from two sources was never in my mind, and my two investigators represented two mutually exclusive possibilities. Perhaps it may help if I describe the precise context in which the trouble arose.

I wanted to make some quantitative assessment of what was lost by introducing a new factor into an experiment, with a view to setting this against the gains. Suppose I have four blocks of 8 plots and I plan a 2^3 experiment on factors A, B, C. Then each of 7 main effects and interactions will be estimated, each with variance $\sigma^2/8$, and σ^2 itself will be estimated by s^2 with 21 d.f. I now want to judge the pros and cons of confounding in a fourth factor, D, on the same plots and blocks. If I do so, I shall presumably still have the same true σ^2, and shall estimate 14 (15 less one confounded) effects with variance $\sigma^2/8$, but s^2, the estimate of σ^2, will now have only 14 d.f. How do I compare the quality of the estimation of, say, the main effect of A by the alternative designs?

It seems to me that I can ask four questions about the information that the experiment provides on the main effect of A:

(1) How much information will the experiment give, if I know the value of σ^2 exactly?

Answer: $8/\sigma^2$.

(2) Having performed the experiment and obtained an estimate, s^2, of σ^2, with n d.f. (or indeed, having such an s^2 from any other source), what is the estimate of the total information that would become available if σ^2 were known?

Answer: $8(n-2)/(ns^2)$, an unbiased estimator of $8/\sigma^2$.

(3) Having performed the experiment as in (2), what is the amount of information that it has given directly, without reference to any other knowledge on σ^2?

Answer: $8(n+1)/\{(n+3)s^2\}$.

(4) What is my expectation under (3) at the start of the experiment? I do not mind this involving an unknown σ^2, as I want it only for comparison of alternative designs *with the same unknown* σ^2.

Answer: Apparently the expectation of (3), evaluated over the distribution of s^2, and therefore $8n(n+1)/\{(n-2)(n+3)\sigma^2\}$.

But (4) is greater than (1), which seems to conflict with common sense!

Moreover, if two designs allow n_1, n_2 d.f. for estimates of variance, the efficiency of the second relative to the first in respect of estimating the main effect of one factor would be, from (3)

$$\frac{(n_1+3)(n_2+1)}{(n_1+1)(n_2+3)}\,\frac{s_1^2}{s_2^2}.$$

If the second design in fact involves taking (n_1-n_2) contrasts from the first error and assigning them to effects and interaction of new factors, so that

$$n_1s_1^2 = n_2s_2^2 + (n_1-n_2)s_3^2,$$

where s_2^2, s_3^2 are independent mean squares estimating σ^2, the expectation of the efficiency can be integrated out to

$$\frac{n_2(n_2+1)(n_1-2)(n_1+3)}{n_1(n_1+1)(n_2-2)(n_2+3)}$$

which agrees with (4) above. My numerical example had $n_1=21$, $n_2=14$, whence the efficiency works out at 190/187: thus the 2^4 confounded design appears to be better for estimating the main effect of one of the original factors, quite apart from its merits in respect of extra factors. Of course this can also be seen because my result in (4) *decreases* monotonically to its limit $8/\sigma^2$ as n increases. This again seems to be at variance with common sense.

Now, where do I run off the rails?

[P.S.] I see no reason why you should spend time in putting me right, but I thought that the paradox might amuse you.

Fisher to D.J. Finney: 24 *March* 1955

Thank you for your letter. I do not think anyone would impute malice to you, but, of course, you are not the only one who has referred to these terms. In the process of claiming for Neyman that which he borrowed from me, Pearson, page 75 of *Biometrika Tables*, says:

> 'The fiducial theory of R.A. Fisher and the confidence interval theory of J. Neyman were developed to meet this situation. Both aim at providing means of calculating from the data intervals within which the unknown parameter θ may be expected to lie within a given measure of probability. While from the practical standpoint there is in most (though not in all) problems no difference between the result of applying the two methods, Fisher's approach introduces the concept of fiducial probability, while Neyman's employs only the classical concept of direct probability. The application of theory described here is in terms of Neyman's approach.'

The description in Pearson's words is as false as it can be, since in both the argument starts with the same probability statement, and in the end Neyman denies the possibility of making any probability statement about the natural facts behind the observations, while the argument I introduced allows for probability statements being made *a posteriori* in the absence of knowledge *a priori*, such as Bayes' argument requires. Where I was mistaken in 1930 was in not making it clear that the absence of knowledge *a priori* is a pre-requisite for the fiducial argument, so that it is not possible, as I wrongly thought at that time, to have both a Bayesian and a fiducial probability statement in apparent conflict.

In the same preface, page 18:

'It will be seen that the two limits are random variables whose values depend only on the observations; their position and distance apart will vary from one sampling to another. The probability statement must be interpreted in the sense that if we determine the limits for σ from the observations by the formulae given above, we are following a procedure which will give in the long run an interval estimate including the true σ in a proportion of about $1 - 2\alpha$ cases. Further, this statement is true even if the successive samples are drawn from populations having different standard deviations. The limits are termed *confidence limits* and $1 - 2\alpha$, often expressed as a percentage, is termed the *confidence coefficient*.'

Pearson takes here a typical example of a fiducial probability statement and takes the trouble to say that its limits, and the probability itself, have the names 'confidence limits' and 'confidence coefficient', so ascribing this novelty of statistical reasoning to himself and Neyman. In this case, however, he clearly differs from Neyman, who would deny that any such probability statement can be made about features of the real world.

I am sorry you did not read my letter further than to give, as you do on page 2, other features which, in your opinion, are inherent in a statement in terms of probability. Here you evidently read a great deal more into the word than I do, for though I said, and though you repeat, the phrase 'the logical content of a probability statement', you seem to want to put into it all that belongs to the framing of a scientific theory. I do not believe that the gambler's belief that the probability of a card chosen at random shall be a king is 1/13 involves more than the two elements, which I was trying to make plain to you, when all these further ideas came into your head, involving what hypotheses should be framed about the distribution of an observed measurement.

As regards your paradox, you ought to consider whether your 'unbiased estimate' gives you satisfactory guidance in the cases of *one* or *two* degrees of freedom.

D.J. Finney to Fisher: 1 *April* 1955

Your two quotations are certainly horrid misrepresentations.

I'm sorry you should think that I did not read your long letter fully, but it is the kind of letter that can be re-read profitably. I think I now see your point more clearly. The situation is more confusing when a continuous variate is involved. For a discrete variate, your two elements suffice to lead to the occurrence of events depending upon a multinomial distribution, or something of that kind.

For a continuous variate (e.g. a random sample of 10 Englishmen measured for height), the nature of the inference that can be expressed probabilistically depends upon the knowledge that we have about the set of possibilities and their relative frequencies. If we know nothing about these except that definite frequencies exist, we are presumably thrown back on rather restricted

non-parametric or distribution-free inferences about the population, but this complies with your requirements. If we know that individual heights are normally distributed, that constitutes different information on the possible values and their frequencies. I see now that what I was doing in my earlier letter was to try to list different levels of knowledge about these possibilities, without realizing that all were covered by the first of your two 'elements'. I was not concerned with framing a theory, but with trying to particularize the knowledge with which we start: I now see this to be unnecessary from the point of view of general discussion of necessary and sufficient conditions for making probability statements, though it may become more important in particular instances when we want to know what statements of probability can justifiably be made from certain data.

Meanwhile, my paradox remains with me. I should of course have qualified my previous discussion with '$n>2$'. I think I am right in saying that, for n degrees of freedom,

$$E(1/s^2) = n/\{(n-2)\sigma^2\} \text{ for } n > 2,$$

the integral failing to converge for $n = 1$ or $n = 2$. I don't see why one should not admit the impossibility of estimating $1/\sigma^2$ for an experiment that allowed only 2 d.f. for s^2, but the apparent inconsistencies for larger n remain.

Fisher to D.J. Finney: 18 *April* 1955

I only mentioned the case of $n = 1$ or 2 in order to raise the question in your mind as to the relevance of the expectation of $1/s^2$. I won't admit at all that one knows nothing about an unknown variance when one has one or two degrees of freedom from which to estimate it; in fact I should assert that one can derive in either of these cases an explicit fiducial distribution of the unknown variance.

With Barnard's help I have been wrestling with the relationship between axiomatic assertion of ignorance, inductive reasoning, uncertain inference, etc., and in case you are interested I enclose two little sections from two different chapters of the book I am putting together.

D.J. Finney to Fisher: 6 *May* 1955

I have not replied earlier to your letter of 18 April because I wanted a chance to read your enclosures two or three times. Although I find my original problem something of a mystery still, your own ideas are most stimulating. The preparation of your new book is excellent news. I have been hoping for something like this for years, and these four pages excite my interest greatly. Very many thanks for letting me see them.

Fisher to D.J. Finney: 17 October 1960

. . . I have been working on the distribution of the weighted mean of small normal samples. Like Behrens with an extra parameter. If the normal deviate x defines the level of significance as usual then the correction will be

$$\begin{array}{l} \text{a cubic in } x \text{ and } d \div n_1 \\ \text{5th power in } x \text{ and } d \div n_1^2 \text{ and } n_1 n_2 \\ \text{7th power terms} \div n_1^3, n_1^2 n_2. \end{array}$$

These last are very troublesome to clean up.

The thing is almost too complicated to be useful, but I should like to see how it tabulates. My impression is that d (Sukhatme's Test) is more important than one would guess, so that approximations taking $d = 0$, or even $d = 1$, would be pretty bad. . . .

D.J. Finney to Fisher: 25 October 1960

. . . Your work on the weighted mean sounds interesting. I have in the past used the standard Fisher–Behrens' tables to give something approximating to fiducial statements about the weighted mean of two quantities where each weight is based upon an ordinary variance estimate. I have described this on page 31 of my book on biological assay, though I do not think that the idea had much originality. I think that you once pointed out to me that this was not a true fiducial argument because, on the hypothesis that the two statistics estimate the same parameter, some information on variance is contained in the difference between them.[1] Am I right in thinking that it is this extra piece of information that you are taking into account in your present study? My guess would be that, unless the degrees of freedom for the two samples are exceedingly small, the information contained in the differences of the two means would be negligible. It will be interesting to compare results based on your new distribution and table with those from the approximation that neglects this information. . . .

[1] See Fisher's letter of 2 December 1949 (p. 90).

Fisher to D.J. Finney: 1 November 1960

A few years ago I came to the conclusion that no one would be able to judge of the value of some approximate approach to the small sample weighted mean problem until he could compare with exact values. Clearly it involves n_1, n_2, θ, d and P, and the practical question is: How important is d?
Using the set-up,

$$\begin{aligned} \bar{x} &= (s_2^2\bar{x}_1 + s_1^2\bar{x}_2)/\{s_1^2 + s_2^2\}, \\ 1/S^2 &= (1/s_1^2) + (1/s_2^2), \\ \mu &= \bar{x} + Su, \end{aligned}$$

then if

$$P = \int_x^\infty e^{-\frac{1}{2}v^2}\,dv,$$

$$u = x + \Sigma\{u_{pq}/n_1^p n_2^q\}$$

and the first approximation is given by

$$u_{10} = \tfrac{1}{4}\{(x^3 + x)c^4 + 4(x^2 + 1)dc^3s + 2x(3d^2 - 1)c^2s^2 + 4d(d^2 - 1)cs^3\}$$

and u_{qp} may be obtained from u_{pq} by interchanging c and s and reversing d.

I do not believe that we need more than the first term to judge that d is rather important.

Payne gave the adjustments for n_1^{-2} and $n_1^{-1}n_2^{-1}$ though I do not know how far he completed this job. I am trying to perfect the terms for (3) and (21)[1] . . .

[1] cf. *CP* 285.

Fisher to D.J. Finney: 3 April 1962

. . . Have you followed at all the activities of David Sprott of Waterloo, and D.A.S. Fraser of Toronto? They have recently sent me a very ingenious form for the correlation coefficient. Namely,

$$\frac{r}{\sqrt{1-r^2}}\,\chi_{N-2} - \frac{\rho}{\sqrt{1-\rho^2}}\,\chi_{N-1},$$

a pivotal function with a standardized normal distribution, involving the two random variables χ. Choosing any value of ρ, it yields the sampling distribution of r, as I gave it in 1915 [*CP* 4], while if you choose any value of r, you can obtain the distribution of ρ as I gave it in 1930 [*CP* 84], and in some new form such as,

$$\frac{1}{\pi(N-3)!}\quad(1-r^2)^{\frac{1}{2}(N-2)}\,(1-\rho^2)^{\frac{1}{2}(N-3)}\,d\rho\quad\left\{\frac{\partial}{\sin\theta\,\partial\theta}\right\}^{N-3}\frac{\theta-\frac{1}{2}\sin 2\theta}{\sin^3\theta}.$$

They have both been inclined to throw Fiducial methods in the teeth of the Americans; who indeed have been pretty dumb about it. There will probably be an ISI session about Inference at Ottawa next year . . .

16

D.J. Finney to Fisher: mid April 1962

. . . I have seen a little of the work that Sprott and Fraser have been doing, but I had not seen this extraordinarily interesting proposal for the correlation coefficient. It looks as though they are in the process of doing some very valuable work by putting the formal mathematics of the fiducial argument in a way that will compel the attention of some of those who have previously disregarded it. Am I right, though, in thinking that their treatment of fiducial inference is at present limited to certain classes of distributions and parameters and does not embrace all the uses that you have made? I got the impression from their first paper that they had not subsequently read more, and I am not sure how generally they expect to be able to produce suitable pivotal functions. . . .

Fisher to D.J. Finney: 3 *May* 1962

. . . I do not suppose that Rao, Fraser and Sprott have cracked the inversion problem, but I suppose the direct simultaneous distribution will appear in *Sankhyā* shortly. I hope also the RSS may soon get around, after more than a year, to publishing my paper on Bayes' experimental procedure [*CP* 289].

I think Fraser must have started with some rather over-simple notions in which the phrase 'continuous group of transformations' was prominent, and exhaustiveness and monotonicity not very visible. Probably he had given little attention to Estimation. Still he extricated himself from the Tukey–Savage sort of rubbish. I do not know whether he has made any *amenda* about the correlation coefficient, in print at least.

My chief intention in the last example in *Statistical Inference* was to illustrate the existence of 3 strata of parameters and their corresponding statistical estimates. This rather complicates the exact statement of the conditions for simultaneous fiducial distributions. I believe I said something about it in my last letter, but I forget how much.

Click here for next section